Evolution of the knowledge base in knowledge intensive sectors

Jackie Krafft, Francesco Quatraro, Pier-Paolo Saviotti

Evolution Of The Knowledge Base In Knowledge Intensive Sectors

Jackie Krafft*, Francesco Quatraro**, Paolo Saviotti***

* University of Nice Sophia Antipolis, CNRS GREDEG, and OFCE DRIC
** University of Nice Sophia Antipolis, CNRS GREDEG, and University of Torino
*** INRA GAEL, and University of Nice Sophia Antipolis, CNRS GREDEG

1. Introduction[1]

In a knowledge based society the creation and utilisation of knowledge become the key factors determining the competitiveness of firms, regions and countries. In this perspective a considerable effort is today dedicated to characterise the knowledge base of different sectors in the economy and to detect its impact on firm performance and on industrial organization (Breschi, Lissoni, and Malerba, 2003; Krafft, 2004; Nesta and Saviotti, 2005). Although all sectors in modern economies are affected by a growing knowledge intensity, some sectors are influenced more than the average. We call these Knowledge Intensive Sectors (KISs). In this paper we map the dynamics of knowledge generation within three KISs: biotechnology, telecommunications and electronics. The first question which is addressed is how to characterize a KIS. Typically we would expect KISs to have a high R&D intensity, to produce more patents and publications than less knowledge intensive sectors and to have a greater impact of knowledge production on firm performance and on sectoral growth. A further and important aspect of KISs is the presence of discontinuity in knowledge. Not that such discontinuities are present only in KISs: other sectors are going to be affected, although often less directly, by these discontinuities. However, KISs are likely to be the first ones to start exploring new forms of knowledge and to move them towards exploitation. Thus, we can expect the dynamics of knowledge generation and utilization in KISs to be affected by both (i) the rate of knowledge creation and (ii) the presence of discontinuities in new knowledge. It follows that in order to be able to link the dynamics of knowledge creation and utilization to firm performance and to industrial organization we need to detect a number of properties of the knowledge base (KB) of KISs. Properties such as the diversity/variety of the KB, its coherence and its cognitive distance (or conversely its similarity) between different KBs have already been shown to be potential determinants of firm performance. The aim of this paper is to contribute to this new literature by characterizing the evolution of the KB in three KISs, namely biotechnology, telecommunications and electronics. We use data from the European Patent Office database (EPO database) to see whether we can find common trends in the evolution of the KB of these three KISs.

The approach developed in this paper is original in that it considers the sector level and not the firm level. A number of contributions have centred so far on the issues of

[1] This work is part of a research project funded by Agence Nationale de la Recherche (contract number: ANR JCJC06_141306).

corporate diversification and corporate coherence (Piscitello, 2004; Nesta, 2007), and have thus a direct background in the economics of the firm (with essentially Penrose, 1959; and Teece, Rumelt, Dosi, and Winter, 1994, as key references). We investigate a different, though complementary question. To the extent that the performance of each sector is driven by knowledge and that each KIS has a different KB we expect specific patterns and trends to emerge at the sectoral level. Of course, this does not mean that firm specificity disappears within sectors. On the contrary, the study of knowledge dynamics at the level of the firm constitutes a complementary question which deserves a separate treatment. We analyse the changes in the composition of the KBs of the three KISs, describing each one of them by means of a set of technological fields found in the International Patent Classification (IPC classes).

To improve our knowledge of the nature of KISs and of the processes of their evolution our methodology will follow two steps, and will proceed to test four research propositions.

The first step consists of documenting the high rate of knowledge production in KISs. Basic statistics like patent counts, the share of patent applications within each sector and the growth trend in the dynamics of patent applications will be mobilized. All these statistics show that the rate of knowledge production in KISs is high, although with a larger distribution of patents over classes for electronics compared to biotechnologies and telecommunications.

The second step relates to the knowledge discontinuities that characterize the KISs. In previous contributions (Grebel, Krafft and Saviotti, 2006; Saviotti, 1996) we hypothesized that when sectors are faced with the emergence of a radically new type of knowledge corresponding to a new paradigm, random search strategies generally precede more organized search strategies. If this hypothesis holds we expect to observe the following regularities in the evolution of the properties of the KB of KISs. First, during the random screening period we would expect to observe (i) an increase in the variety of the technological classes associated with each KIS, (ii) a fall in the coherence of the KB of KISs, (iii) a growing cognitive distance between the internal KB of incumbent firms previously used in the KIS and the new emerging knowledge. On the other hand during the organized screening period we would expect the rate of growth of variety to fall, coherence to rise, and cognitive distance to decline. The results we will show later seem to generally confirm these predictions. However, important differences can also be identified among the KISs, especially for what concerns the respective length or the persistence over time of the random *versus* the organized period.

On the basis of the previous considerations we can now formulate the following four propositions:

P1: A knowledge intensive sector has a higher than average rate of knowledge creation.
P2: The emergence of a discontinuity in a type of knowledge suitable to become the future knowledge base of a KIS leads to the sequence of the two periods of random search first and of organized search later.
P3: During the random search period KB variety rises, KB coherence falls and the cognitive distance between the previous KB of a KIS and the new emerging

knowledge rises. During the organized search period KB the rate of growth of variety falls, KB coherence rises and the cognitive distance between the previous KB of a KIS and the new emerging knowledge falls.

P4: The higher the rate of increase over time in variety and in cognitive distance, and the higher the decrease over time in coherence in the knowledge base, the more persistent the period of random screening, i.e. the less established the organized screening period.

In the rest of the paper Section 2 presents the data. Section 3 discusses the basic statistics about the dynamics of knowledge production in KISs, and analyses P1. Section 4 explores the existence of a phase of random search, followed by a phase of more organised search for each KISs, and discusses P2. Section 5 defines the measure of the knowledge base as a set of components including variety, coherence, cognitive distance, and other complementary indicators, and elaborates on P3. Section 6 generates the empirical results from the application of these measures to the database and discusses P4. Section 7 concludes.

2. The Data

The initial dataset of patent applications consists of 2,659,301 items, both EU and Worldwide applications, over the period 1978 – 2005. Our search strategy is based on queries reporting the IPC classes that define each KISs under study, namely biotechnology, telecommunications and electronics. The analysis thus focuses on three subsets of patent applications, identified by merging the classifications set up by the OECD and the *Observatoire des Sciences et des Techniques*. We adopted these classifications to elaborate some rough boundaries to our KISs, although we will realize later in this paper that in some cases these classifications leave some important classes out.

Taking into account these elements, it resulted that the biotechnology sector includes 11 IPC classes, the telecommunications sector is made up of 16 IPC classes and the electronics sector consists of 30 IPC classes (see Table 1).

Table 1 – Definition of sectors using IPC classes

BIOTECHNOLOGY	
A01H	new plants or processes for obtaining them; plant reproduction by tissue culture techniques
A61K	preparations for medical, dental, or toilet purposes
C02F	treatment of water, waste water, sewage, or sludge
C07G	compounds of unknown constitution
C07K	peptides
C12M	apparatus for enzymology or microbiology
C12N	micro-organisms or enzymes; compositions thereof
C12P	fermentation or enzyme-using processes to synthesise a desired chemical compound or composition or to separate optical isomers from a racemic mixture
C12Q	measuring or testing processes involving enzymes or micro-organisms; compositions or test papers thererof; processes of preparing such compositions; condition-responsive control in microbiological or enzymological processes

C12S	processes using enzymes or micro-organisms to liberate, separate or purify a pre-existing compound or; processes using enzymes or micro-organisms to treat textiles or to clean solid surfaces of materials
G01N	investigating or analysing materials by determining their chemical or physical properties
TELECOMMUNICATIONS	
G08C	transmission systems for measured values, control or similar signals
H01P	waveguides; resonators, lines, or other devices of the waveguide type
H01Q	aerials
H03B	generation of oscillations, directly or by frequency-changing, by circuits employing active elements which operate in a non-switching manner; generation of noise by such circuits
H03C	modulation
H03D	demodulation or transference of modulation from one carrier to another
H03H	impedance networks, e.g. resonant circuits; resonators
H03K	pulse technique
H03L	automatic control, starting, synchronisation, or stabilisation of generators of electronic oscillations or pulses
H03M	coding, decoding or code conversion, in general
H04B	transmission
H04H	broadcast communication
H04J	multiplex communication
H04K	secret communication; jamming of communication
H04L	transmission of digital information, e.g. telegraphic communication
H04Q	selecting
ELECTRONICS	
F21H	incandescent mantles; other incandescent bodies heated by combustion
F21K	light sources not otherwise provided for
F21L	lighting devices or systems thereof, being portable or specially adapted for transportation
F21M	transferred to F21s and F21V
F21P	transferred to F21s and F21V
F21Q	transferred to F21s and F21V
F21S	non-portable lighting devices or systems thereof
F21V	functional features or details of lighting devices or systems thereof; structural combinations of lighting devices with other articles, not otherwise provided for
G05F	systems for regulating electric or magnetic variables
H01B	cables; conductors; insulators; selection of materials for their conductive, insulating, or dielectric properties
H01C	resistors
H01F	magnets; inductances; transformers; selection of materials for their magnetic properties
H01H	electric switches; relays; selectors; emergency protective devices
H01J	electric discharge tubes or discharge lamps
H01K	electric incandescent lamps
H01M	processes or means, e.g. batteries, for the direct conversion of chemical energy into electrical energy
H01R	electrically-conductive connections; structural associations of a plurality of mutually-insulated electrical connecting elements; coupling devices; current collectors
H01T	spark gaps; overvoltage arresters using spark gaps; sparking plugs; corona

	devices; generating ions to be introduced into non-enclosed gases
H02B	boards, substations, or switching arrangements for the supply or distribution of electric power
H02G	installation of electric cables or lines, or of combined optical and electric cables or lines
H02H	emergency protective circuit arrangements
H02J	circuit arrangements or systems for supplying or distributing electric power; systems for storing electric energy
H02K	dynamo-electric machines
H02M	apparatus for conversion between ac and ac, between ac and dc, or between dc and dc, and for use with mains or similar power supply systems; conversion of dc or ac input power into surge output power; control or regulation thereof
H02P	control or regulation of electric motors, generators, or dynamo-electric converters; controlling transformers, reactors or choke coils
H04M	telephonic communication
H05B	electric heating; electric lighting not otherwise provided for
H05C	electric circuits or apparatus specially designed for use in equipment for killing, stunning, enclosing or guiding living beings
H05F	static electricity; naturally-occurring electricity
H05K	printed circuits; casings or constructional details of electric apparatus; manufacture of assemblages of electrical components
Source: World Intellectual Property Organization.	

Table 2 reports the count and the share of each sector within the whole dataset in terms of patent applications. Although the biotechnology sector is defined by the lowest number of IPC classes, its share in the overall dataset is the highest (12.08%), while the sector gathering the highest number of classes, i.e. electronics, represents the lower share (1.81%). The telecommunications sector occupies an intermediate position in both the number of IPC and its share in the dataset.

Table 2 – Overall distribution of the three sectors

	#	%
Biotechnology	321449	12.08
Telecommunications	115735	4.35
Electronics	47955	1.81

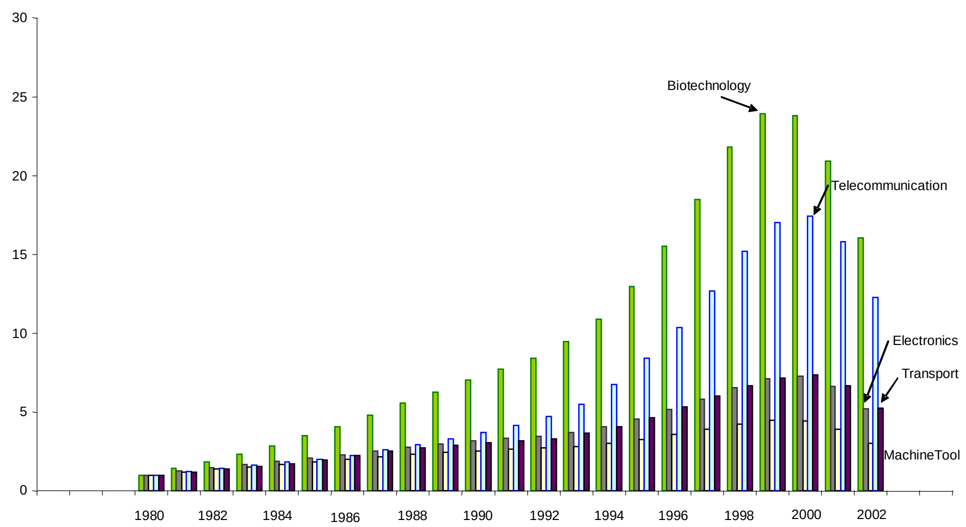
3. Dynamics of knowledge production in KISs

One can reasonably assume that the dynamics of knowledge production in KISs is marked by important specificities. In this perspective, one immediate (and potentially obvious) specificity is that knowledge production in KISs is likely to be higher than in other sectors. In this section, we explore whether the following hypothesis "the higher knowledge production, the more knowledge intensive the sector" holds in our KISs (namely, biotechnology, telecommunications and electronics).

3.1. Increasing growth trend

A closer look at the dynamics of patent applications reveals that all of the three sectors are characterized by an exponential growth trend. This can be explained by the general increase in the propensity to patent which is often attributed to the diffusion of European patents among innovating agents and the increase of resources devoted to R&D activities (Archontopoulos, E., Guellec, D., Stevnsborg, N., Van Pottelsberghe de la Potterie, B., Van Zeebroeck, N., 2007). However, in our case, a series of specific features appears concerning the KISs under study. Biotechnology, telecommunications and electronics have a higher rate of knowledge production than other sectors such as machine tools or transport (see Figure 1).

Figure 1: Index of the (5-year-)moving average of the number of patents with the base year 1980
(source: Grebel, Krafft, Saviotti, 2006)



Although at a first glance our KISs seem to share a common growth pattern, the behaviour of biotechnology is quite different from that of telecommunications and of electronics. The number of patents in biotechnology is about twice as large as in telecommunications and about three times as large as in electronics. Furthermore, the rate of growth of patents in biotechnology seems to be more evenly distributed during the period studied than in telecommunications.

3.2. Distributional properties of classes and co-occurrences

Figures 2, 3 and 4 report the dynamics of technological classes in each sector. The number of IPC classes can be interpreted as a rough measure of the differentiation, or scope, of each sector's knowledge base. First of all it is immediately clear that the dynamics of technological differentiation in the three KISs is influenced by the dynamics of the patent stock. The rate of growth in the number of IPC classes closely parallels the rate of growth of the number of patents. Indeed all of the three sectors show an increasing number of technologies, the growth trend being exponential in the case of biotechnology and telecommunications, and linear in the case of electronics.

It is interesting to note that the stock effect influences the time trend of the variable, but not the cross-sector differences. Indeed the electronics sector is characterized by a higher number of technological classes than biotechnology or telecommunications, although its yearly number of patent applications is sensibly lower. In addition to the time path of the total number of technological classes also their distribution within each KIS deserves to be analyzed. It turns out that in each of the three KISs few classes occur very frequently while most other classes are very rarely used. As a consequence we only used for our statistics the top 10 classes in order of frequency of occurrence according to their relative share in 1992. This means that some classes may appear or disappear from the subset we use to represent the knowledge base of the KIS as a result of the changing composition of the knowledge base. In fact this changing composition is an expression of the structural change which occurs in science and which gives rise to structural change in output.

Figure 2 – Patent Classes in the Biotechnology sample of patents, by year

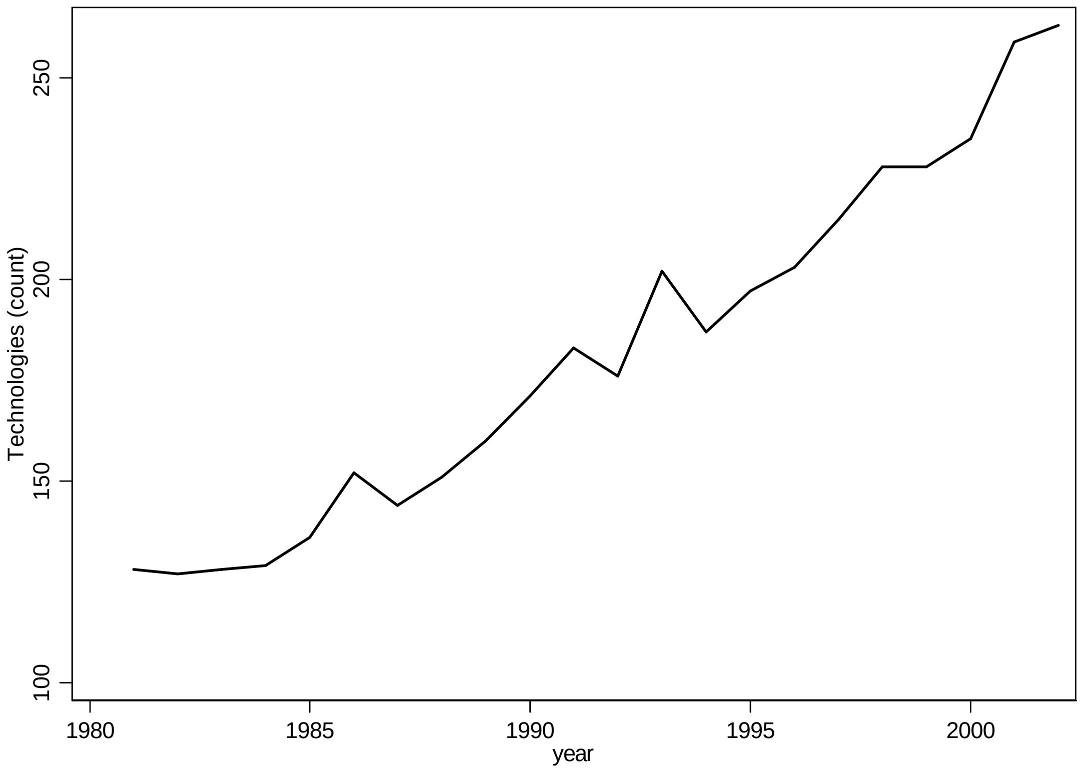


Figure 3 - Patent Classes in the Telecommunications sample of patents, by year

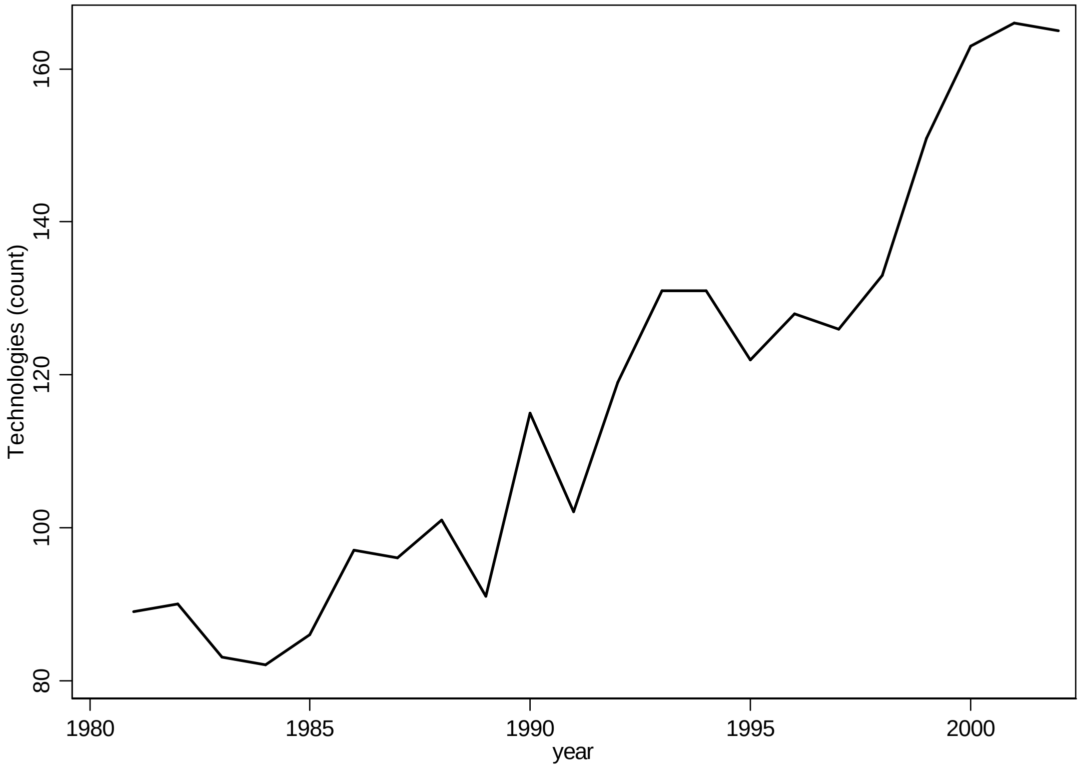


Figure 4 - Patent Classes in the Electronics sample of patents, by year

A number of interesting considerations stem from Table 3, where the shares of IPCs in the biotechnology sector are reported. First of all, the **A61K** class retains the highest share all over the observed period. It refers to a very generic set of industrial applications, ranging from cosmetics to medicinal preparations containing antigens or antibodies. Secondly, some IPCs appear in Table 3, while they were not listed in Table 1. This suggests that the boundaries of sectors taken from OECD and OST classifications are only fuzzy boundaries that need to be complemented by the detailed observation of data. For example, C07C and C07D, which are typical chemical classes and which were used in the pharmaceutical industry before the emergence of modern biotechnology, remain components of the KB of pharmaceutical firms although their share of total patents falls over time. Conversely, **A61P**, which appears in 1988 and whose share dramatically increases in 2000, is a class which while not included amongst the descriptors of biotechnology covers medicinal and therapeutic applications of chemical compounds. The last class, together with **C12N** and **C07K**, seem to be the most interesting in our subset in that they are characterized by marked growth dynamics. Finally, some classes such as **C12Q**, **C02F** and **C12P** while being distinctly biological show either a less pronounced growth or even a fall in their relative weight over the time period considered. These classes correspond to more 'traditional', and hence less fast growing, applications of biotechnology, such as fermentation techniques. The above considerations remind us that defining the boundaries of knowledge types (disciplines, fields etc) is not easier than defining the boundaries of industrial sectors.

Table 3 – Dynamics of Top 10 IPCs share in Biotechnology, by year

	A61K	C07D	C12P	C12N	C07K	A61P	C12Q	C07H	C02F	C07C
1981	0.360	0.161	0.041	0.032	0.001		0.018	0.023	0.054	0.086
1982	0.366	0.158	0.039	0.039	0.001		0.021	0.023	0.051	0.085
1983	0.367	0.146	0.044	0.053	0.006		0.021	0.027	0.039	0.083
1984	0.360	0.130	0.056	0.072	0.035		0.020	0.024	0.041	0.058
1985	0.343	0.113	0.068	0.080	0.065		0.022	0.023	0.029	0.045
1986	0.338	0.107	0.065	0.087	0.075		0.027	0.025	0.036	0.037
1987	0.363	0.105	0.059	0.088	0.071		0.025	0.024	0.033	0.032
1988	0.354	0.095	0.066	0.094	0.079	0.000	0.027	0.032	0.029	0.033
1989	0.344	0.091	0.059	0.101	0.082	0.000	0.035	0.030	0.030	0.032
1990	0.339	0.085	0.060	0.114	0.085	0.001	0.037	0.028	0.031	0.027
1991	0.339	0.087	0.057	0.109	0.094	0.001	0.044	0.032	0.029	0.028
1992	0.341	0.091	0.057	0.104	0.091	0.002	0.046	0.030	0.028	0.027
1993	0.344	0.087	0.051	0.111	0.089	0.002	0.045	0.029	0.027	0.028
1994	0.347	0.085	0.041	0.116	0.097	0.003	0.048	0.030	0.027	0.026
1995	0.344	0.080	0.035	0.127	0.105	0.003	0.050	0.032	0.025	0.025
1996	0.341	0.077	0.037	0.124	0.102	0.003	0.054	0.030	0.023	0.023
1997	0.339	0.074	0.030	0.128	0.107	0.002	0.061	0.029	0.023	0.020
1998	0.332	0.067	0.029	0.143	0.110	0.002	0.070	0.024	0.020	0.020
1999	0.316	0.062	0.028	0.140	0.109	0.040	0.065	0.028	0.020	0.020
2000	0.290	0.060	0.025	0.126	0.101	0.116	0.054	0.023	0.018	0.016
2001	0.289	0.057	0.021	0.123	0.111	0.120	0.059	0.019	0.016	0.016
Source: our elaborations on EPO data.										

In telecommunications (Table 4) the change in the weights of the leading IPCs is even more marked than in biotechnology. In 1981 the IPC with the highest share was **H03K**. The share of this class decreased so rapidly that at the beginning of the 2000s it was no longer within the leading group. **H04Q** showed a soft decrease in the 1980s followed by a pretty fast rate of growth in the second half of the 1990s, thus becoming the second ranking IPC in telecommunications, overtaking the class **H04B**. The latter class is indeed characterized by a modest increase in the 1980s, and then by slowdown in the second half of the 1990s. Unsurprisingly, the class that emerges as the leader in the second half of the 1990s is **H04L**, which is related to digital transmission. This class already outperforms H04B in 1996 and takes persistently the leadership until 2001.

Table 4 – Dynamics of Top 10 IPCs share in Telecommunications, by year

	H04B	H04L	H03K	H04Q	H01Q	H04J	H01P	H03H	H03M	H04M
1981	0.125	0.119	0.139	0.104	0.051	0.032	0.030	0.034		0.038
1982	0.131	0.123	0.127	0.074	0.055	0.037	0.035	0.029		0.035
1983	0.111	0.115	0.164	0.092	0.051	0.042	0.033	0.030	0.001	0.029
1984	0.105	0.120	0.152	0.069	0.055	0.031	0.034	0.038	0.026	0.020
1985	0.129	0.126	0.132	0.063	0.052	0.031	0.020	0.038	0.048	0.021
1986	0.119	0.110	0.130	0.080	0.053	0.038	0.037	0.033	0.042	0.036
1987	0.130	0.130	0.112	0.057	0.088	0.026	0.040	0.030	0.052	0.022
1988	0.131	0.125	0.101	0.085	0.053	0.031	0.036	0.042	0.049	0.020
1989	0.120	0.142	0.099	0.069	0.071	0.045	0.042	0.033	0.046	0.025
1990	0.156	0.129	0.097	0.085	0.059	0.045	0.021	0.038	0.050	0.020
1991	0.163	0.140	0.082	0.076	0.051	0.051	0.032	0.033	0.052	0.020
1992	0.159	0.147	0.083	0.078	0.057	0.050	0.039	0.037	0.028	0.024
1993	0.180	0.143	0.076	0.116	0.045	0.042	0.020	0.033	0.034	0.038
1994	0.170	0.163	0.062	0.116	0.050	0.047	0.022	0.036	0.042	0.030
1995	0.159	0.153	0.066	0.146	0.044	0.051	0.023	0.028	0.043	0.031
1996	0.153	0.180	0.057	0.158	0.049	0.048	0.023	0.033	0.037	0.030
1997	0.158	0.188	0.046	0.163	0.047	0.048	0.021	0.022	0.035	0.044
1998	0.159	0.198	0.037	0.184	0.050	0.047	0.016	0.018	0.030	0.040
1999	0.167	0.214	0.030	0.190	0.045	0.051	0.015	0.014	0.033	0.043
2000	0.163	0.230	0.030	0.175	0.043	0.048	0.017	0.015	0.028	0.039
2001	0.150	0.244	0.023	0.170	0.046	0.044	0.018	0.017	0.031	0.037
Source: our elaborations on EPO data.										

Table 5 reports the share of the top 10 classes in the electronics sectors. Looking at the figures, and comparing them with the evidence about the other two sectors, the idea that here patents are more distributed over the classes gains further support. Indeed no class has a share exceeding 9% over the whole period, and the lowest share is 1%. This means that the rest of the classes observed in the electronics sector retains a weight lower than 1%. Bearing this in mind one class, namely **H04M,** emerges as the leader in the second half of the 1990s. Clearly, this is a class closely related to the telecommunications sector. For this reason it is hardly surprising that its share increases over time, particularly in the late 1990s. The rest of the classes stay in between the 2% and 3%, the only exception being **H05K**, which reaches 6% in the 1990s. Even in this case, the evidence is not surprising, as patents in this class refer, among the others, to printed circuits (i.e. assemblies of individual semiconductors), which are of paramount importance in the manufacturing of electronic devices used for digital communication and data storing. The growing importance of H04M and of H05K shows that telecommunications is one of the main sectors of application of electronics. However, the relatively flat distribution of IPC classes shows that electronics has a wider range of applications to other sectors than either biotechnology or telecommunications, both of which are more internally focused.

Table 5 - Dynamics of Top 10 IPCs share in Electronics, by year

	H05K	H01B	H01J	H05B	H01R	H04M	H01F	H02G	H01H	H01L
1981	0.034	0.046	0.026	0.049	0.033	0.036	0.039	0.027	0.026	0.023
1982	0.030	0.054	0.029	0.064	0.028	0.037	0.029	0.019	0.028	0.020
1983	0.046	0.056	0.023	0.048	0.037	0.035	0.033	0.011	0.034	0.021
1984	0.045	0.056	0.044	0.053	0.036	0.019	0.044	0.019	0.036	0.026
1985	0.049	0.050	0.036	0.049	0.035	0.024	0.036	0.018	0.027	0.020
1986	0.051	0.049	0.031	0.052	0.033	0.034	0.027	0.016	0.034	0.011
1987	0.051	0.042	0.050	0.034	0.043	0.026	0.040	0.018	0.032	0.021
1988	0.052	0.047	0.040	0.045	0.033	0.022	0.035	0.023	0.031	0.026
1989	0.048	0.038	0.036	0.050	0.046	0.026	0.041	0.026	0.030	0.023
1990	0.054	0.049	0.039	0.049	0.038	0.028	0.031	0.026	0.025	0.022
1991	0.057	0.051	0.038	0.053	0.049	0.029	0.026	0.033	0.023	0.026
1992	0.061	0.051	0.044	0.042	0.041	0.032	0.031	0.028	0.027	0.027
1993	0.050	0.038	0.034	0.042	0.036	0.051	0.027	0.024	0.023	0.022
1994	0.056	0.037	0.030	0.042	0.051	0.051	0.027	0.026	0.026	0.020
1995	0.049	0.030	0.040	0.048	0.046	0.051	0.032	0.023	0.026	0.024
1996	0.060	0.028	0.039	0.038	0.036	0.057	0.035	0.022	0.025	0.029
1997	0.060	0.028	0.037	0.033	0.035	0.071	0.030	0.021	0.020	0.028
1998	0.049	0.028	0.032	0.039	0.036	0.072	0.028	0.020	0.023	0.027
1999	0.044	0.028	0.034	0.032	0.028	0.090	0.022	0.020	0.021	0.025
2000	0.056	0.023	0.036	0.034	0.027	0.090	0.019	0.016	0.019	0.032
2001	0.056	0.025	0.040	0.037	0.026	0.092	0.022	0.018	0.019	0.032
Source: our elaborations on EPO data.										

4. Random search *versus* organized search

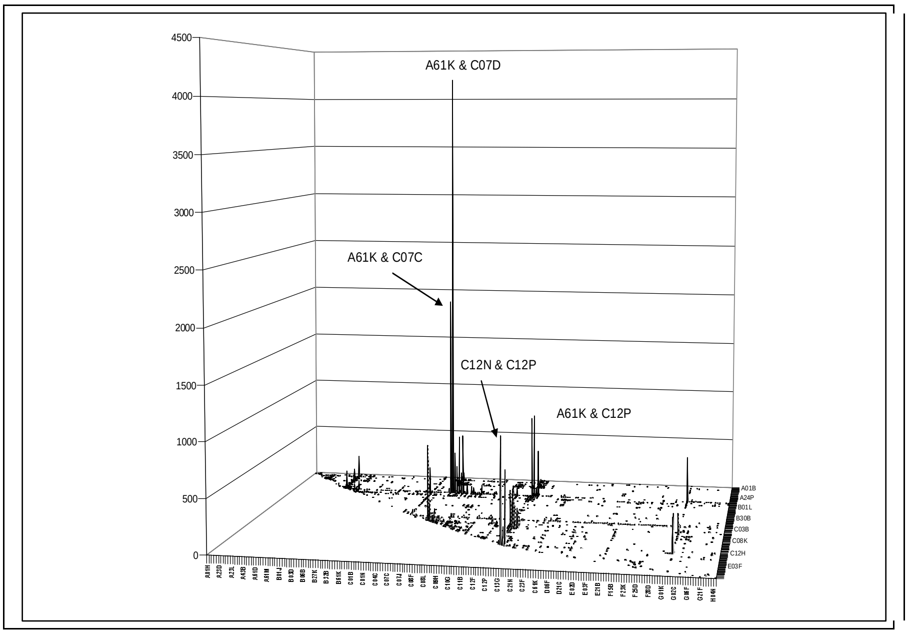
4.1. Patterns of IPC co-occurrences

The evidence presented so far shows that KISs are more knowledge intensive than 'traditional' sectors. Furthermore, there are differences in patenting activity both across and within the three macro-sectors. In a previous paper (Grebel et al, 2006) we had formulated propositions P2-P4 based on the observations of matrices of technological co-occurrence. Such matrices are constructed by representing the IPC classes describing a given type of knowledge on both axes of the matrix. Each patent is classified according to a primary and a number of secondary classes. There is no strict consensus on whether or not this distinction has to be taken into account in investigating the technological base of a sector[2]. In this case we find in the non diagonal cells of the matrix the frequencies of co-occurrence of IPCs. The numerical values of these frequencies are then plotted on a third dimension, thus providing a graphic representation of the distribution of co-occurrences of IPCs in the field of knowledge studied. We interpreted our observation that the distribution of co-occurrences of IPC classes was becoming more uneven as the as the new field of knowledge started maturing as the result of a transition from a random search to an organized search strategy. In what follows we first reconstruct the co-occurrence matrices and then develop a series of measures intended to test more quantitatively propositions P1-P4.

In order to gain a preliminary understanding of the relationships among different classes, and of how they evolve over time, we divided the whole period in 4 sub-periods and summed up the co-occurrences over each period (e.g. the period 1981-1985 presents the cumulated frequency of co-occurrences for each couple of IPCs).

We then obtained square matrices with 0 values on the diagonal by construction and very large dimensions. For this reason only the related diagrams will be shown in what follows. Results are reported in Figures 8 to 16.

Figure 5 - Matrix of classes co-occurence, Biotechnolgy 1981-1985



---

2 In this analysis we follow Breschi et al. (2003), who did not take into account the distinction between primary and secondary classes. See alternatively Verspagen (1997) for the opposite point of view.

Figure 6 - Matrix of classes co-occurence, Biotechnolgy 1986-1990

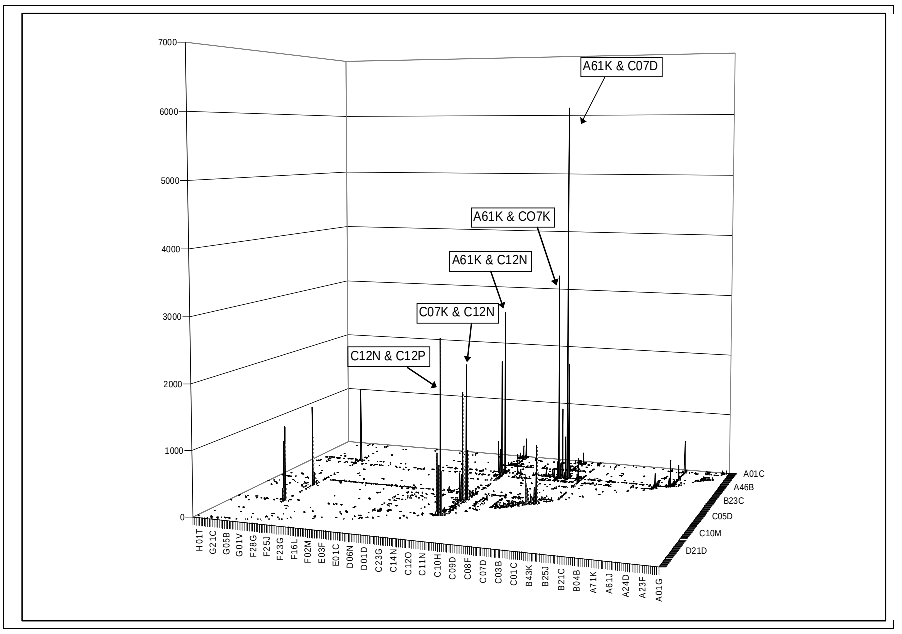


Figure 7 - Matrix of classes co-occurence, Biotechnolgy 1991-1995

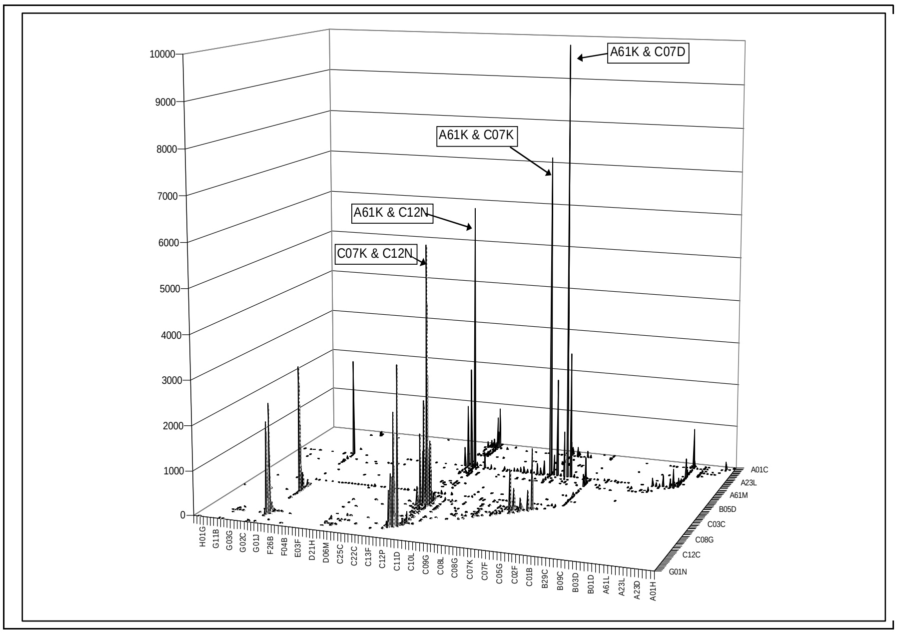


Figure 8 - Matrix of classes co-occurence, Biotechnolgy 1996-2001

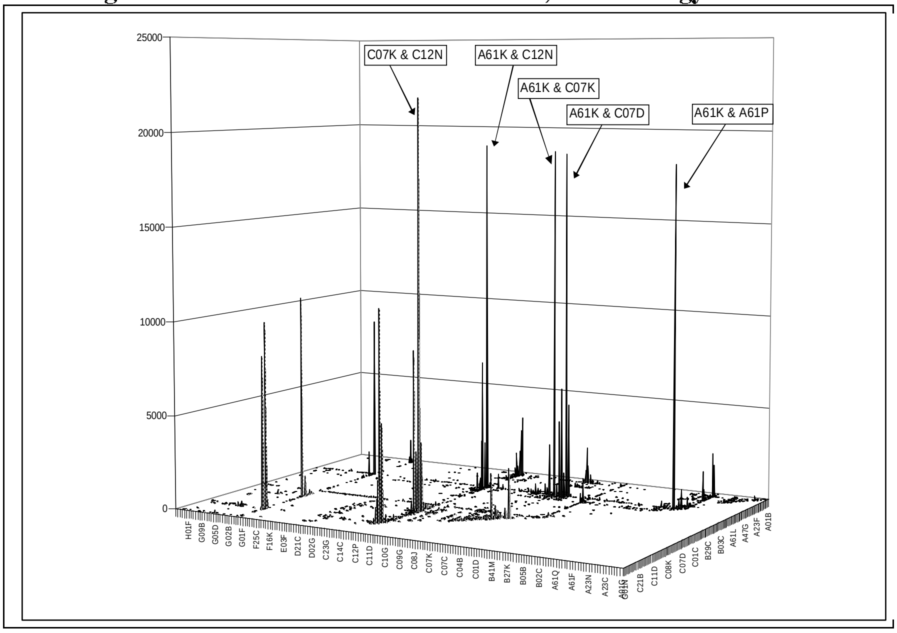


Figure 9 - Matrix of classes co-occurence, TLC 1981-1985

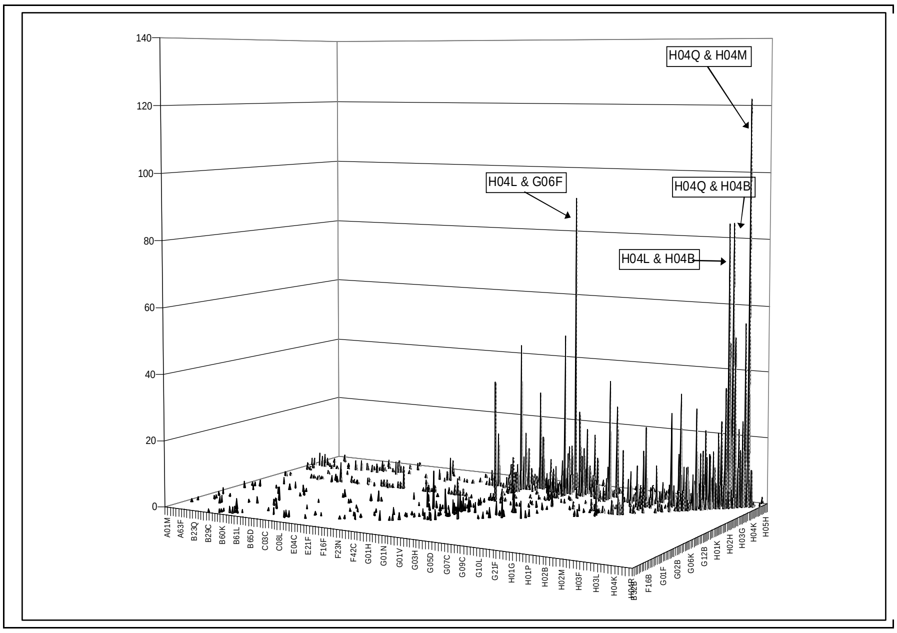


Figure 10 - Matrix of classes co-occurence, TLC 1986-1990

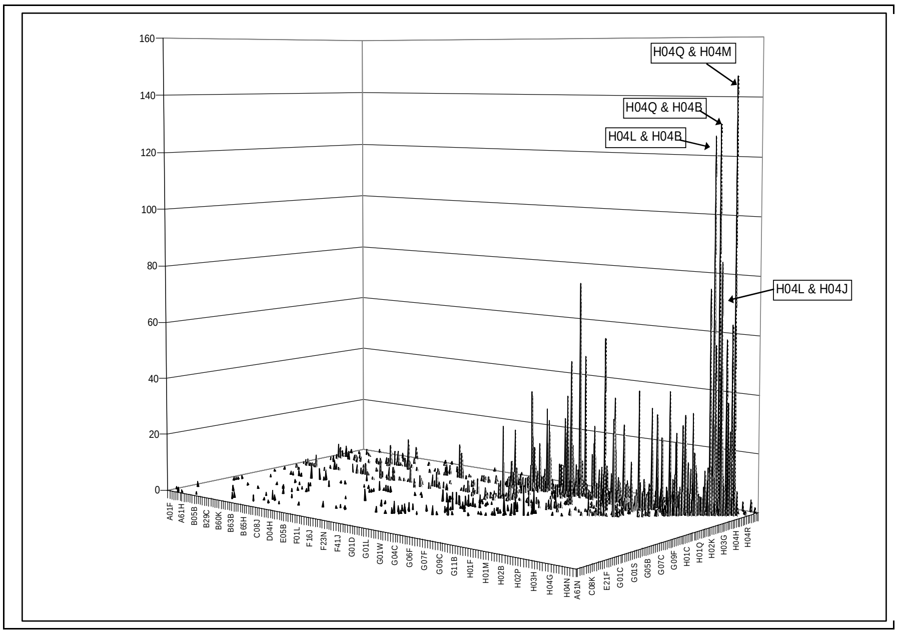


Figure 11 - Matrix of classes co-occurence, TLC 1991-1995

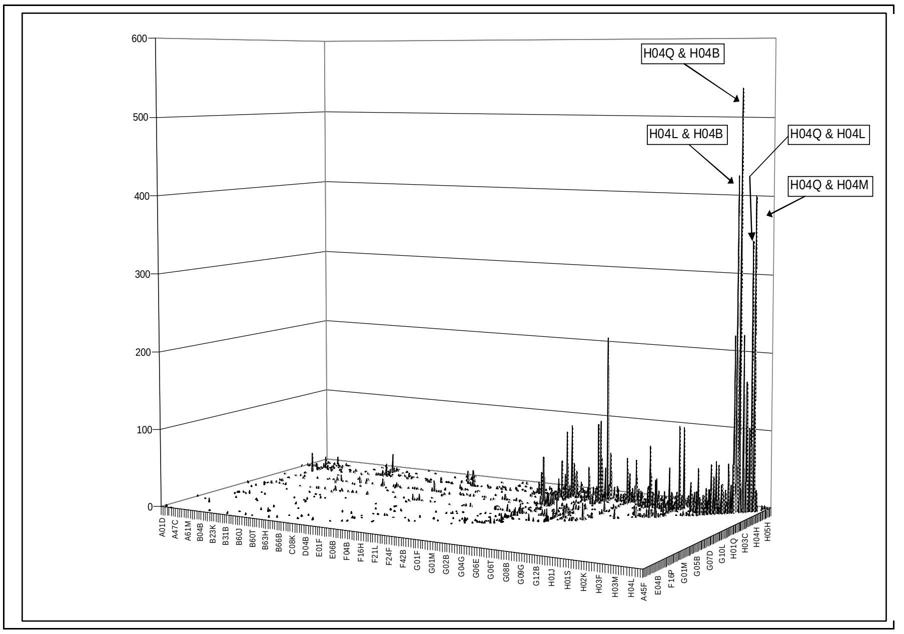


Figure 12 - Matrix of classes co-occurence, TLC 1996-2001

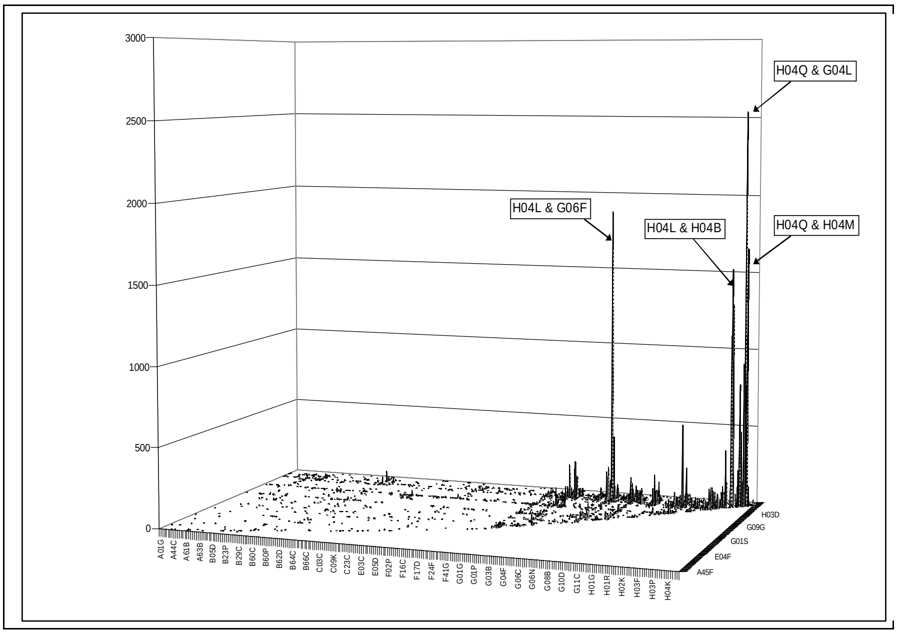


Figure 13 - Matrix of classes co-occurence, Electronics 1981-1985

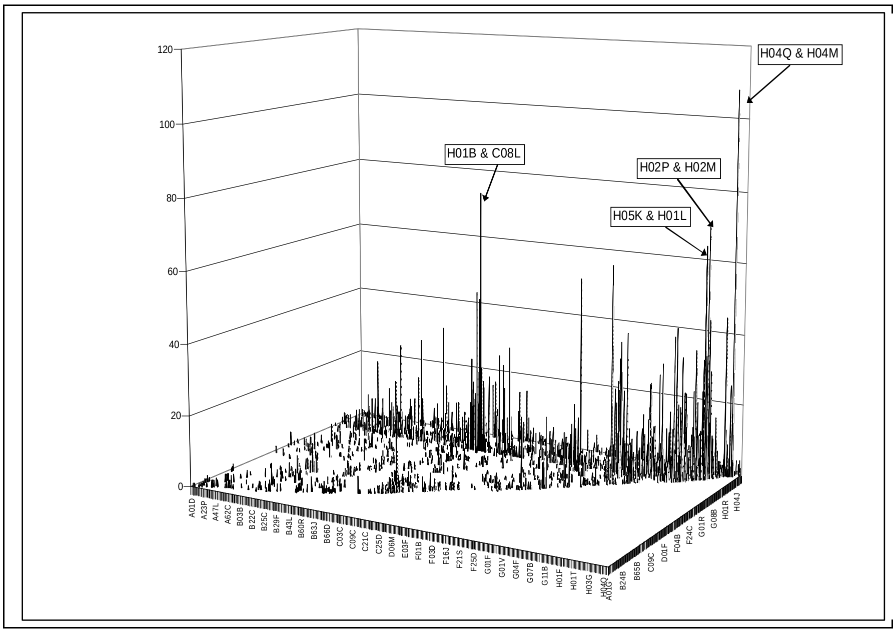


Figure 14 - Matrix of classes co-occurence, Electronics 1986-1990

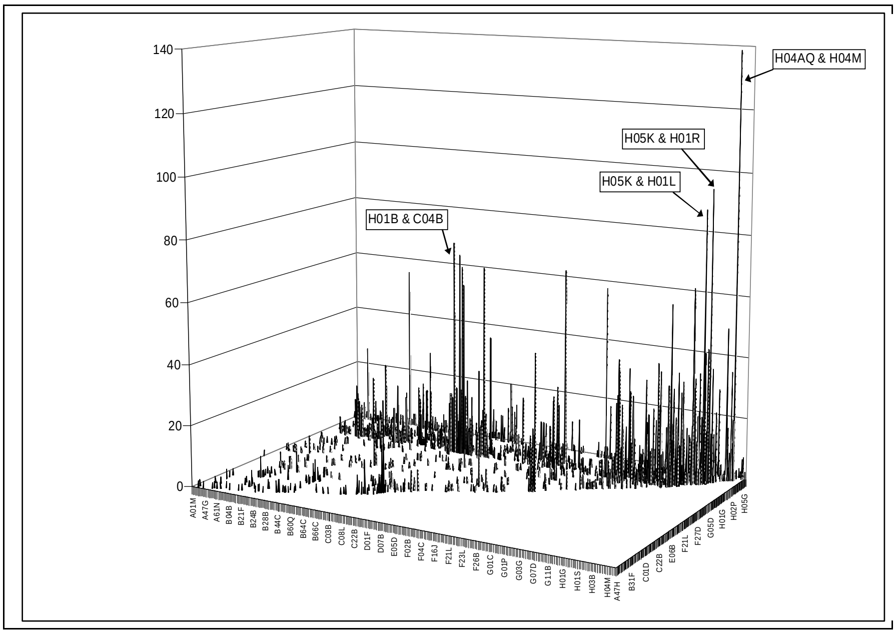


Figure 15 - Matrix of classes co-occurence, Electronics 1991-1995

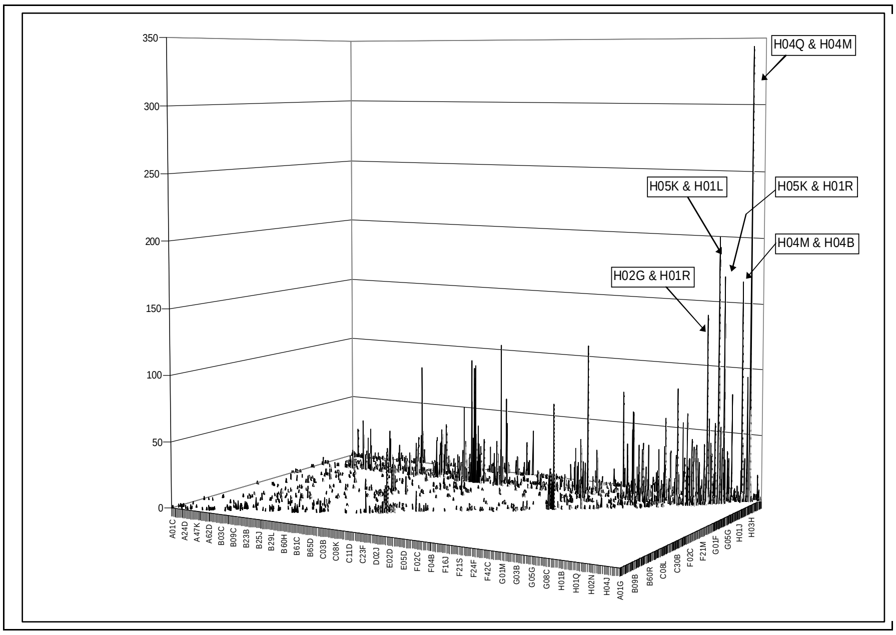
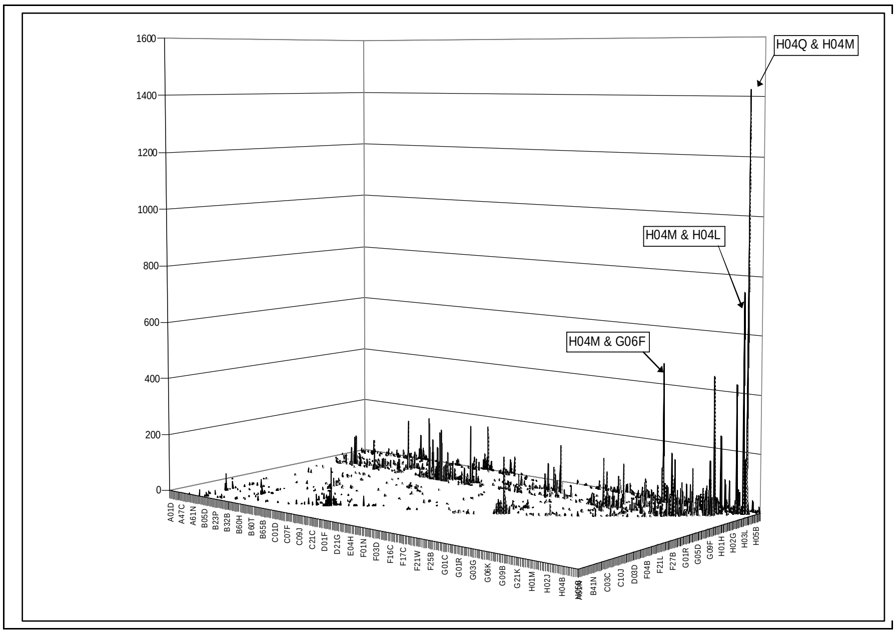


Figure 16- Matrix of classes co-occurence, Electronics 1996-2001



4.2. Common patterns and less common patterns

Our data seem to broadly corroborate proposition P2, or at least they do not contradict it. Search strategies for the new knowledge generally evolve from random search when firms can perceive the opportunities inherent in the new knowledge but have not yet identified promising directions, to a later, more organized search, when most firms can identify within the new knowledge the more promising trajectories. This can be visualized in biotechnologies and telecommunications where, in the earlier periods, a large number of peaks are emerging in Figures 5 to 12 while, in the latter periods, only a smaller number of peaks are observed, also reaching higher values. This common pattern of evolution can be explained as follows. Within the highly uncertain period immediately following the emergence of the new technology the search is random and aimed at learning in all possible directions, stressing differentiation. As the exploration of the new knowledge landscape proceeds, some directions of development emerge as being the most promising. Search becomes more structured around a restricted number of knowledge types and improving the integration of these types of knowledge increases in importance. The shock of novelty produces uncertainty and induces a random search while subsequent learning processes select some subsets of the new knowledge space and structure the search processes around them. The existence of a random search period is an indication that a radical change in knowledge is occurring. The exploration of a completely new part of the knowledge space can be expected to proceed initially without clearly established rules or well defined trajectories.

While the above pattern seems to be compatible with the observed behaviour of biotechnology and telecommunications it seems to fit less well that of electronics. In this case a large number of peaks is observed even in the later periods (see Figure 13 to 16). This different pattern of development is quite likely related to the broader range of applications observed for electronics relative to biotechnology and telecommunications.

5. Measurement of the Knowledge Base

Amongst the various properties of the knowledge base can be measured we select three candidates adapted to our research question: variety, measured by information entropy; coherence, or knowledge relatedness; cognitive distance. We analyse the relationships that may exist between them first during the random search period, and second in the organized search period.

5.1. Variety or information entropy

Entropy measures the degree of disorder or randomness of the system, so that systems characterized by high entropy will also be characterized by a high degree of uncertainty. The greater is the variety within the system, the higher is the amount of information required to describe it, and hence the higher the entropy; the higher the degree of non-equiprobability of states, the lower is the entropy. Lower entropy is also associated with a greater capacity of the system to store information[3]. Interestingly information entropy can also be decomposed in a "within" and a "between" part anytime the events to be investigated can be aggregated in a smaller numbers of subsets. We will use this decomposed index since, in our case, within-entropy will measures variety within the subsets (i.e. within existing or closey related IPC classes), while between-entropy will focus on the subsets measuring the variety across them (i.e., the emergence of new IPC classes). Formally, information entropy is thus defined as the sum of between and average within-group parts (see Frenken, 2008 for details about the calculation of such indexes).

5.2. Coherence or knowledge relatedness

Coherence or knowledge relatedness is a variable aimed at capturing the ability of firms to combine, or integrate, different pieces of knowledge. In fact it is often calculated at the firm rather than at the industry level[4]. At the level of a sector coherence measures the average ability of firms to combine, or integrate, different pieces of knowledge. It is

[3] See Shannon and Weaver (1949) on the initial definition of the entropy index; Theil (1967) on the introduction of such an index in the economic analysis; Saviotti (1986) on the capacity of the system to store information measured by entropy; Attaran (1986), Attaran and Saghafi (1988), Frenken et al. (2007) on applications aimed at measuring diversity of an industry (or of a sample of firms within an industry) against a uniform distribution of economic activities in all sectors, or among firms; and finally Frenken et al. (1999, 2004) on the degree of variety and uncertainty within a technological population.

[4] See Nesta and Saviotti (2005, 2006); Nesta (2007) who defines the coherence of the knowledge base as the average relatedness of any technology randomly chosen within a firm with respect to any other technology. They develop a measure on how much the technologies present within the firm are related each other.

a complementary measure to information entropy. While information entropy analyzes variety from a statistical viewpoint, the coherence index draws upon a relatedness matrix which provides synthetic information on the closeness degree of technologies within the industry portfolio. As an example, increasing entropy signals an increase in the variety of technologies couples observed in the sample of patents. This may be linked to an increasing or decreasing coherence index, according to whether variety concerns related or unrelated technologies. Coherence or (Knowledge Relatedness) has been calculated following Nesta and Saviotti (2006).

5.3. Cognitive distance

This index is intended to measure the dissimilarity between the knowledge bases of different firms, sectors, countries etc. Cognitive distance is defined as the inverse of similarity. Several measures of cognitive distance can in principle be used. The one we use here is derived from the measure of technological proximity originally proposed by Jaffe (1986 and 1989), who investigated the proximity of firms' technological portfolios. Subsequently Breschi et al. (2003) adapted the index in order to measure the proximity, or similarity, between two technologies. Furthermore, this measure is related to Nootebooms' (1999, 2000) concept of cognitive distance.

5.4. Summing up

During the random search period KB variety rises, KB coherence falls and the cognitive distance between the previous KB of a KIS and the new emerging knowledge rises.

During the organized search period KB the rate of growth of variety falls, KB coherence rises and the cognitive distance between the previous KB of a KIS and the new emerging knowledge falls.

6. Empirical Results

6.1. Preliminaries

The measures of information entropy and knowledge relatedness have been applied to investigate the patterns of evolution of knowledge bases in three broad sectors: biotechnology, telecommunication and electronics. As far as entropy is concerned, we considered 4-digit technological classes. 4-digit classes may in turn be assigned to 1-digit larger classes. 1-digit classes will be used to calculate between- and within-group entropy.

The analysis of cognitive distance is carried out by calculating two indexes: the inverse of technological relatedness and the Euclidean distance between two subsequent periods[5]. In what follow the results are indexed with respect to the base year 1981. We

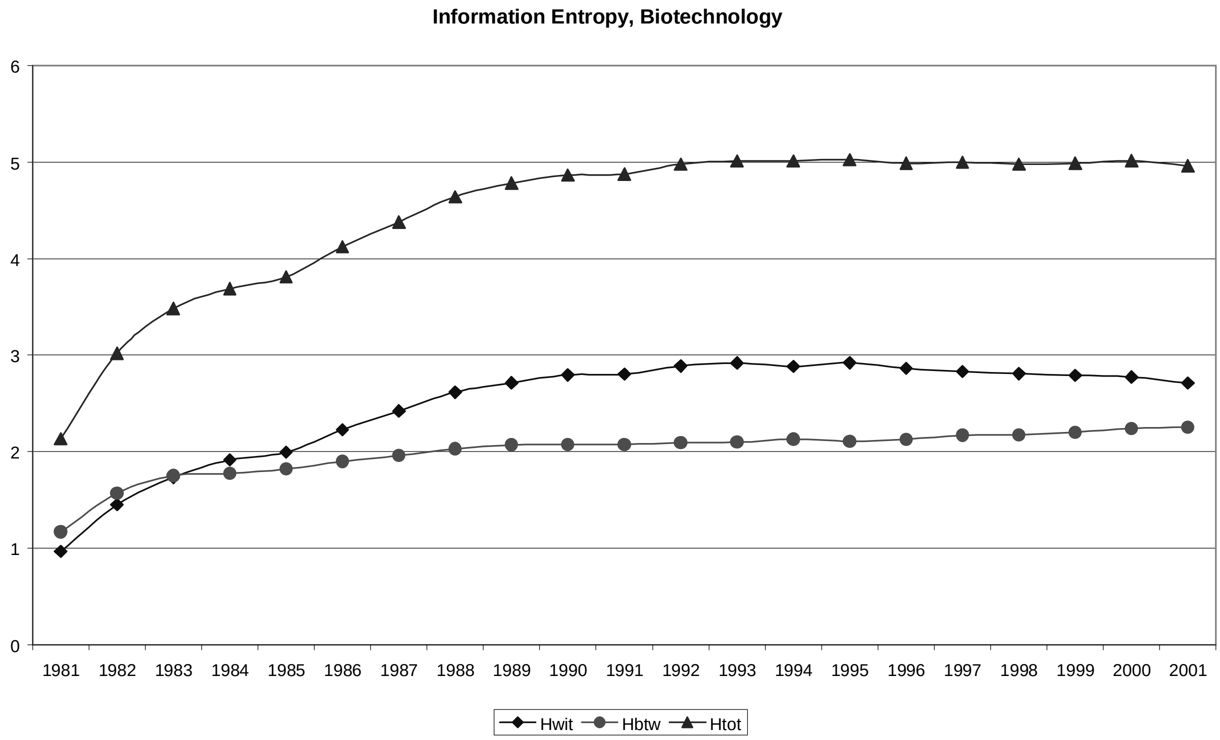
[5] We left out the Euclidean distance between two technologies within the same period because it is conceptually very similar to technological relatedness, but it does not compare the co-occurrence of two technologies *i* and *j* with a third one, with their independent occurrence.

set 1981 = 1 and calculated dC(t)/d*t* at each period. The following years are calculated as C(t+1) = C(t) * (1 + dC(t)/d*t*).

6.2. Biotechnology

Figure 17 shows the evolution of information entropy for the biotechnology sector. The index is positive and above 1 over the whole time period considered. Moreover, the rate of growth of information entropy falls for most of the period of observation until it becomes constant from the early 1990s, with the possible exception of the mid 1980s. In 1985 the rate of growth of variety starts rising in correspondence with the overtaking on inter-group variety by intra- group variety. The distinction of inter- (or unrelated) and intra-group (o related) variety was introduced by Frenken et al. (2007) to measure the output variety of different regions of the Netherlands. In our case while in the early 1980s the between-group entropy was higher than the within-group, the situation was reversed starting from 1985. This would suggest that, while in the very early phases of the emergence of modern biotechnology most of the new knowledge was coming from outside the knowledge base previously used, starting from 1985 internal (to the sector) sources of knowledge differentiation became more prominent.

Figure 17: Information Entropy, Biotechnology

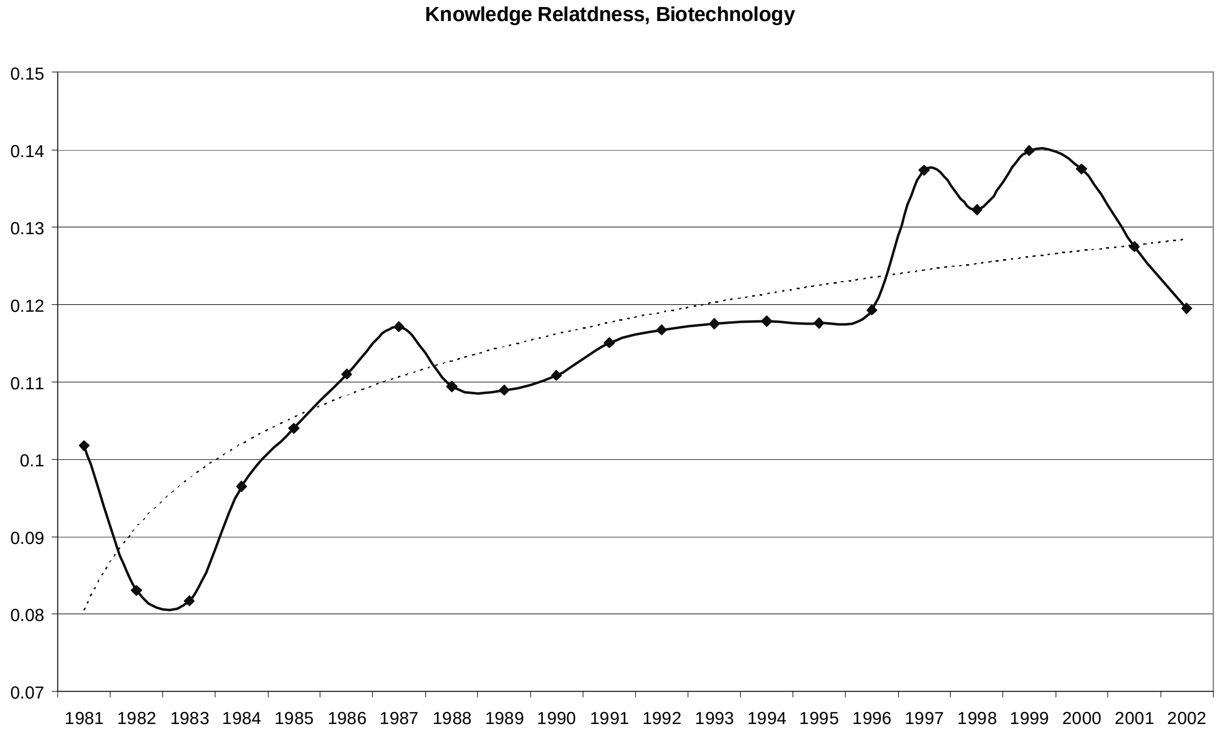


However, the concept of related variety assumes that classes (or sectors) belonging to the same subset are more similar than those which belong to different sets. This may not be the case all the times. As an example consider IPCs A21B and A61K. While they belong to the same subset “A”, they can hardly be said to be technologically related. It may indeed be the case that an increase in the variety of observed technological classes causes an increase in the probability for a technology to co-occur with an unrelated one.

This may be the case also if within-group entropy has higher entropy than between-group. As an example, consider an increase in the co-occurrences of IPC A01B and C12P. It would imply an increase in between-group entropy. However on average their technological relatedness (τ) is 3.99, and hence an increase in their co-occurrences will cause an increase in coherence, other things being equal. If we look at the denomination of the IPCs we find A01B; while C12P. This may well be the case of a patent related to research on chemical fertilizers. So, while the decomposition of entropy helps gaining more information on the processes going on, it needs to be complemented by the coherence index, which allows to assess the technological closeness of classes.

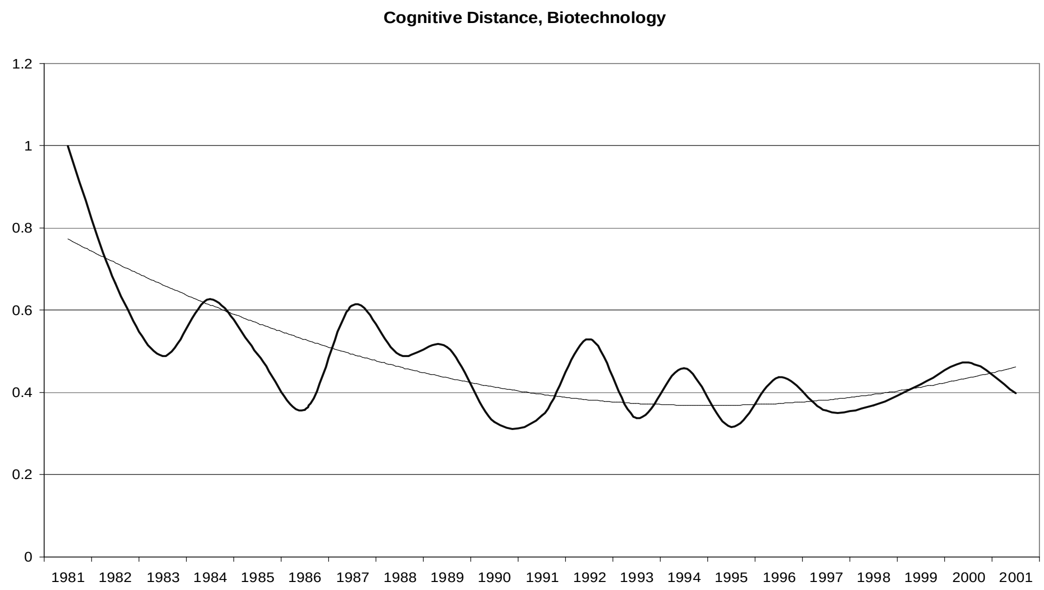
Figure 18 reports the dynamics of knowledge relatedness for the biotechnology sector.
We can distinguish within the overall changes a positive growth trend and superimposed deviations. In particular, there are two periods of fast rise in knowledge relatedness, beginning in 1982 and in 1995 respectively. The first of these deviations from the trend seems to be closely related to the ratio of within-group to between-group variety. When between-group variety is greater than the within-group one, in the period 1981-1982, the coherence index falls. It then begins to increase in 1983 when within-group variety overtakes between-group variety. The subsequent rise in 1997 cannot be explained in the same way. However, it can be observed that the two rises in knowledge relatedness seem to coincide with the onset of the absorption of two different generations of biotechnology, based on recombinant DNA and on genomics respectively, by incumbent firms (Saviotti, Catherine, 2008). Taking this into account we can interpret the overall (trend) rise in knowledge relatedness as due to the relative similarity, or low cognitive distance, of the new types of knowledge which incumbent firms needed to learn. The deviations with respect to the trend could be explained by the emergence of new generations of biotechnology and/or by the ratio of intra to inter group variety. As a new generation of biotechnology emerges the overall trend is not reversed but deviations can occur due to the however limited cognitive distance that the new generation introduces. This line of explanation is not incompatible with the one based on the ratio of within-group to between-group variety. If we assume changes in intra-group variety to be generally lower than those in inter-group variety, then we can expect changes of generation within one technology (e.g. biotechnology) to raise the ratio intra/inter while the emergence of a completely new technology can be expected to lower the same ratio.

Figure 18: Knowledge relatedness, Biotechnology



The evolution of cognitive distance is reported in Figure 19. Even in this case we can distinguish an overall trend from the deviations with respect to it. The evidence for the biotechnology sector is very consistent with the measures of information entropy and knowledge relatedness. The distance index indeed decreases dramatically in the early years of the period we observed. Although with some cyclical fluctuations, it keeps on falling until the first half of the 1990s. Then it remains almost constant with the possibility of a very limited rise.

Figure 19: Cognitive distance, Biotechnology

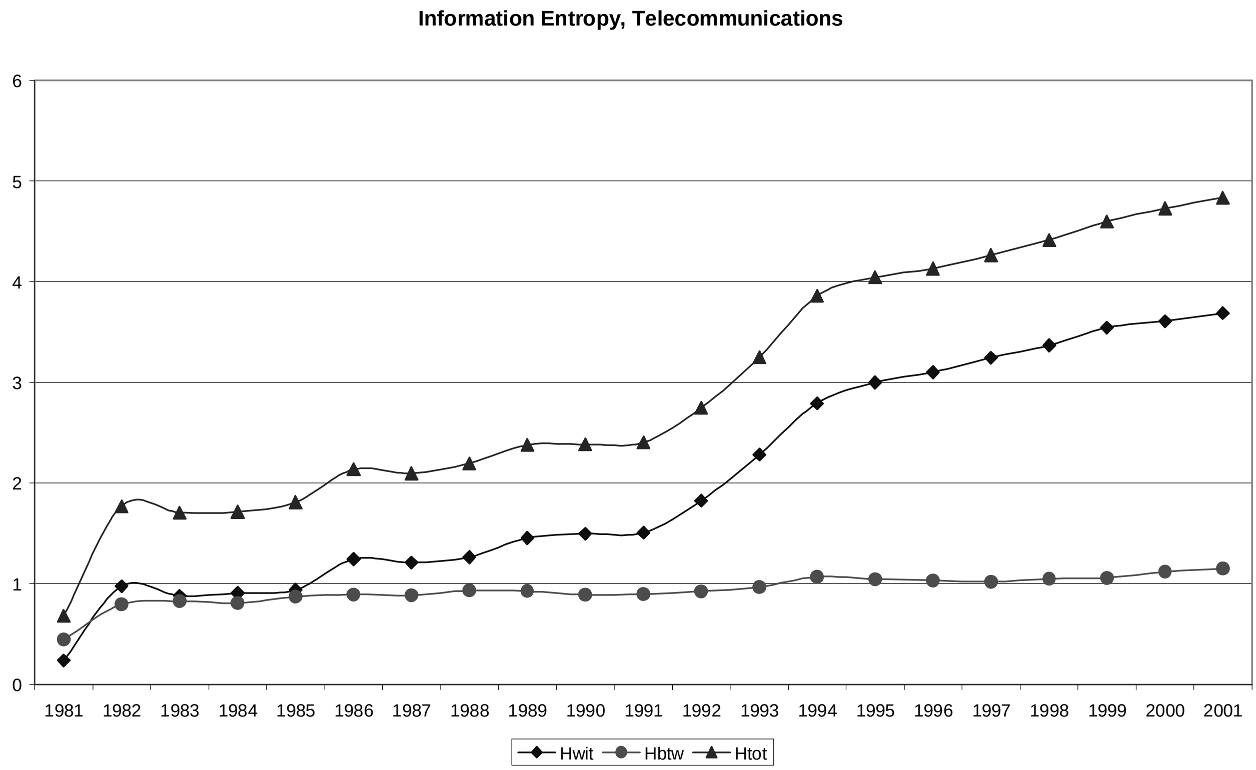


In summary, in the biotechnology sector there has been a growing knowledge differentiation, detected by the growth in variety, accompanied by a trend towards knowledge relatedness and towards falling cognitive distance. These broad trends have been combined with a changing ratio of within to between group variety, with fluctuations with respect to the trend of both knowledge relatedness and of cognitive distance. The overall trend towards a growth in variety implies that the new types of knowledge introduced into the KB of biotechnology are relatively similar to those which were already there. On the other hand, the cognitive distance which these new types of knowledge entailed was enough to cause deviations with respect to the trend. In particular, the deviations were towards a fall in knowledge relatedness at the emergence of a new generation of biotechnology and towards rise in knowledge relatedness as the new generation started maturing. Our findings so far are thus not incompatible with propositions P1-P4. Although the expected fall in knowledge relatedness when the new type of knowledge emerged did not occur, the deviations with respect to the trend bear a close relationship to both the emergence of new generations of biotechnology and to the changing ratio of within- to between-group variety. Propositions P1-P4 were initially formulated without taking into account the distinction between within- and between-group variety and will need to be modified accordingly.

6.3. Telecommunications

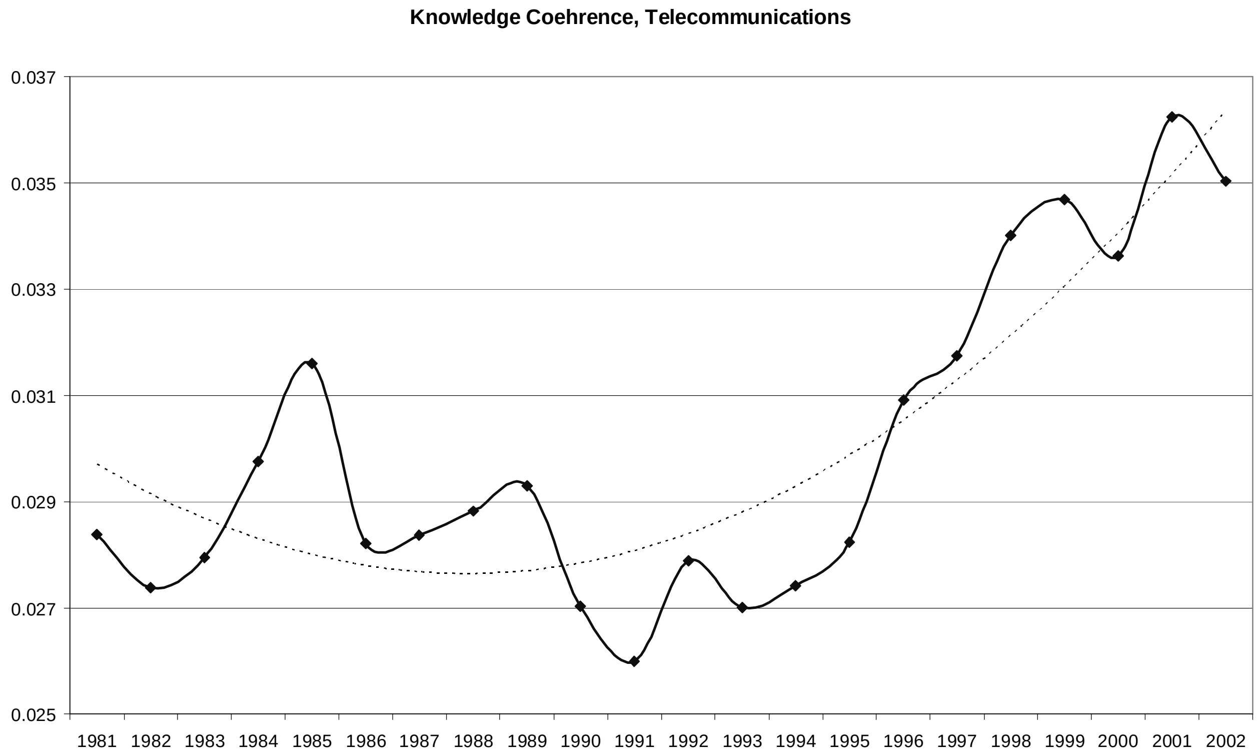
The evidence about Telecommunications is very different from biotechnology. Knowledge variety, as measured by information entropy, increases all the times but it undergoes particularly fast rises two or three times, in 1981, between 1991and 1994 and possibly in 1986 (Figure 20). Furthermore, in this case the between-group entropy is fairly stable all over the period we observed. Hence the dynamics of total entropy is driven by within group entropy. The fastest rise in total variety, occurring from 1991, is totally accounted for by intra-group variety.

Figure 20: Information Entropy, Telecommunications



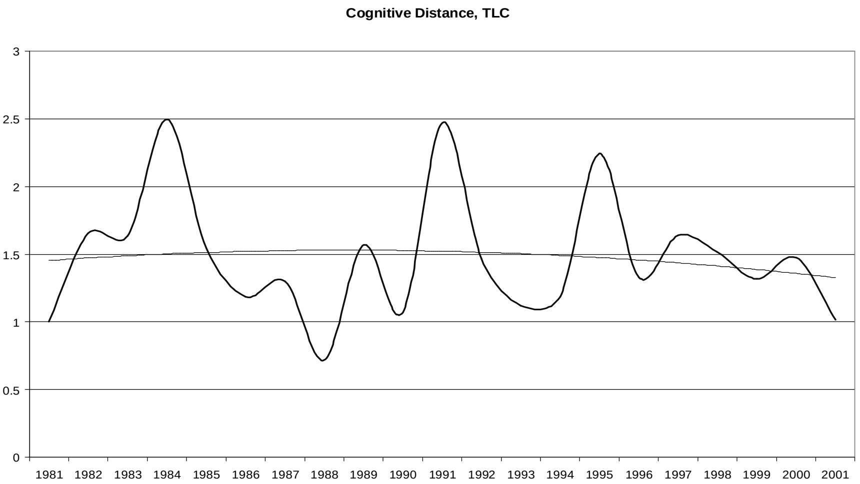
The trend of knowledge relatedness falls until 1990 and then rises (Figure 21). Also in this case the rise in knowledge relatedness seems to be almost completely determined by the growing ratio intra/inter group variety. Even in this case there are large fluctuations with respect to the trend. In particular, the steep increase in the years 1983-1985 is followed by a discontinuity causing a fall in 1986. Then the index keeps falling until 1991 and increases monotonically along the rest of the 1990s. Comparing Figures 20 and 21, one can easily note that the period of steep increase of coherence corresponds to the acceleration of the within group entropy, which occurs especially in the first half of the 1990s. Hence it may be argued that during the 1980s the research efforts in the telecommunication sector were mainly characterized by the search for new technological paths to develop. At mid and at the end of the 1980s two major discontinuities appeared, suggesting the introduction of radical innovations whose technological potential started being explored in the subsequent decade.

Figure 21: Knowledge coherence, Telecommunications



As far as cognitive distance is concerned, the evidence about telecommunications (Figure 22) is less striking than that about biotechnology. Knowledge coherence was slightly decreasing over the 1980s, and then began to increase in the 1990s. Cognitive distance shows a flat trend with a possible slight fall starting in the 1990s. Superimposed upon this trend there are considerable fluctuations.

Figure 22: Cognitive distance, Telecommunications

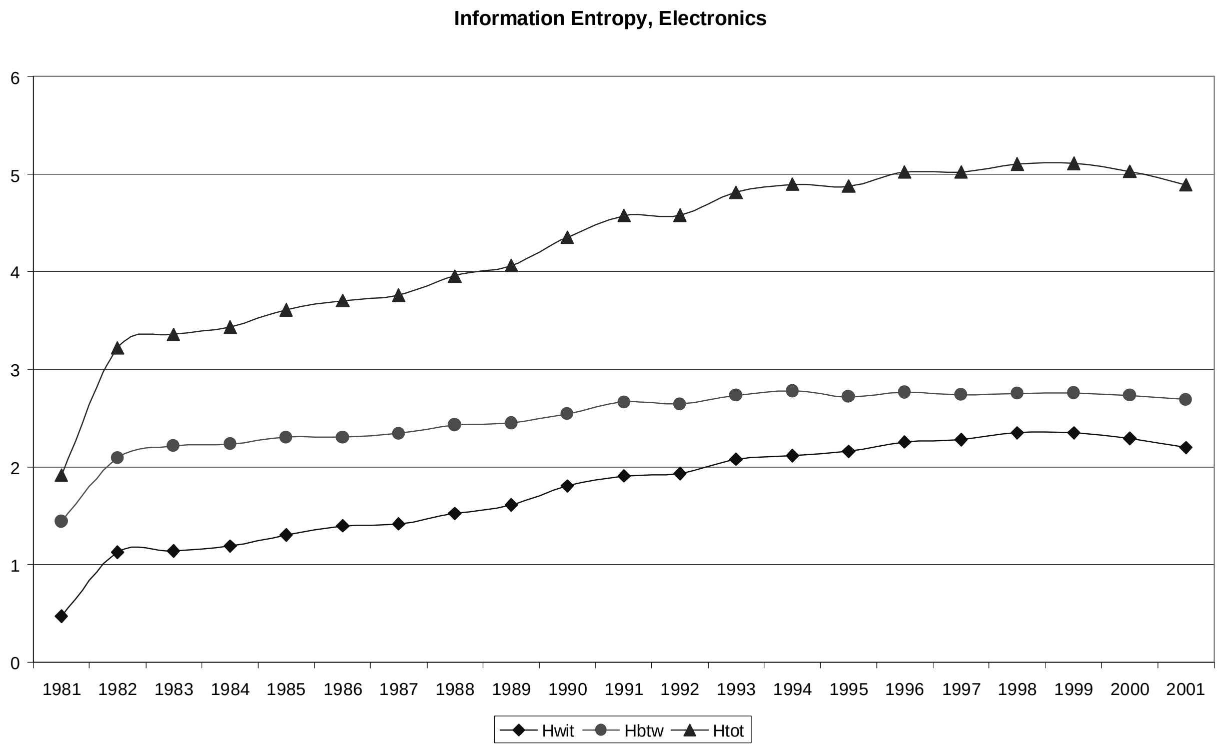


Compared to biotechnology telecommunications shows a less smooth trend towards growing knowledge variety and a larger departure of intra- from inter-group knowledge variety. This indicates that during the period studied new forms of knowledge being used in telecommunications were increasingly similar to those already present within the sector. This interpretation is confirmed by the almost constant value of cognitive distance. Furthermore, the relative rise in intra-group knowledge variety seems to indicate a progressive focusing of new forms of knowledge inside the technology.

6.4. Electronics

In electronics total entropy keeps increasing until 1998, when it basically stabilizes (Figure 23). However, interestingly, in this case total knowledge variety is led by the between-group variety, which always has a higher weight than the within-group one.

Figure 23: Information entropy, Electronics



The trend of knowledge coherence falls all the times (Figure 24). The fluctuations show falls beginning in 1981 and in 1989 and raises beginning in 1985 and in 1993. In this case the overall trend with respect to biotechnology and to telecommunications can be explained by the higher relative value of inter-group knowledge variety. This implies a higher cognitive distance, as confirmed by the diagram in fig 25. The fluctuations around the trend are for the moment more difficult to explain and will require further research.

Figure 24: Knowledge coherence, Electronics

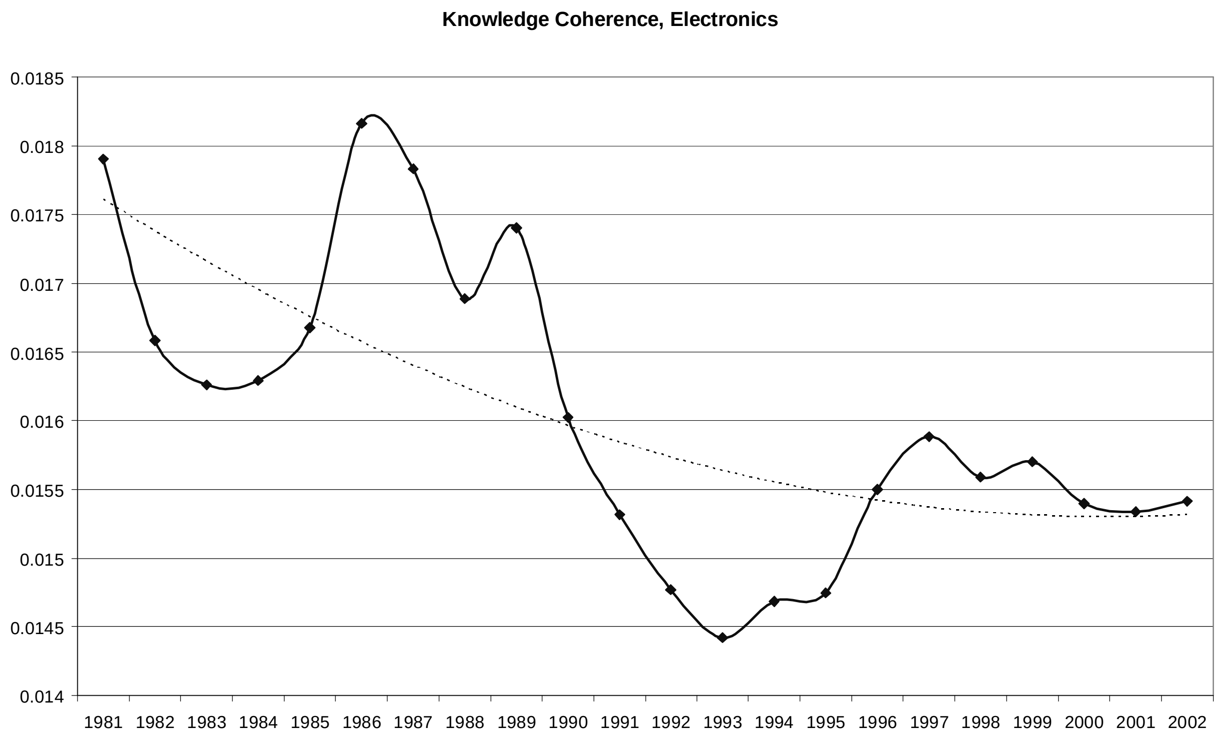
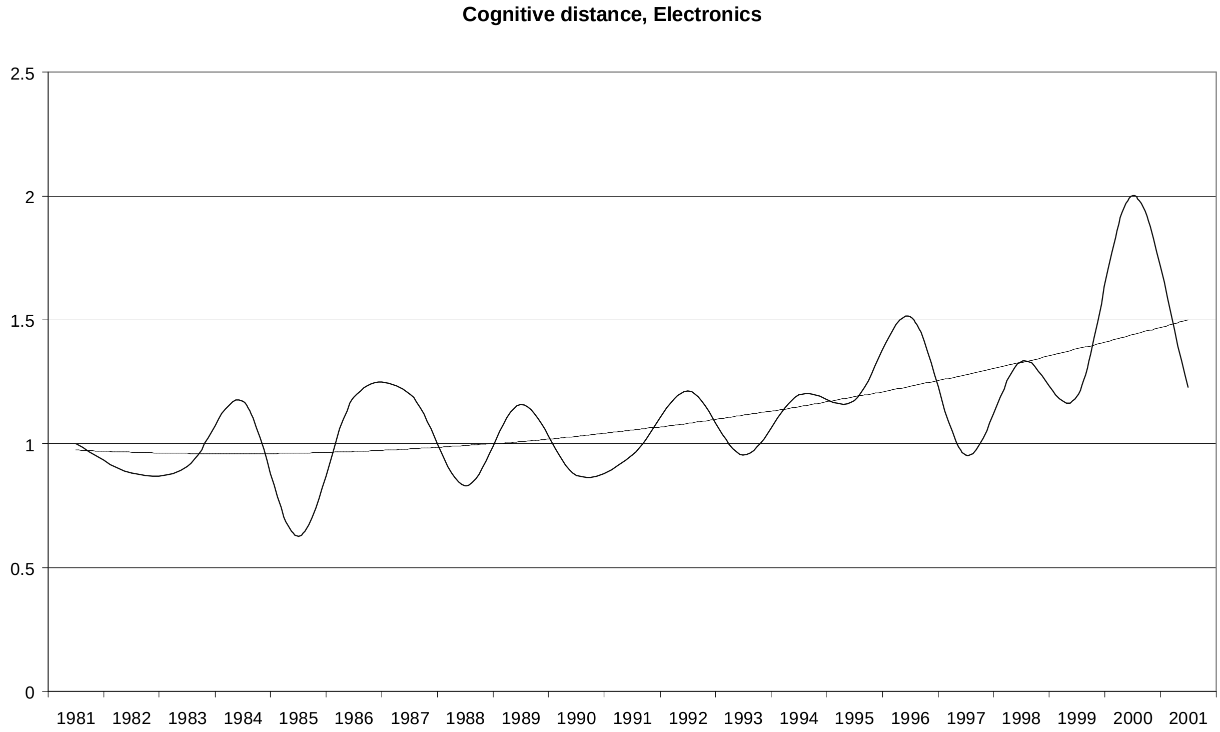


Figure 25: Cognitive distance, Electronics



6. Summing up

Our initial hypothesis that the emergence of a discontinuity in knowledge should lead to the subsequent adoption of a random search strategy first and of an organized search strategy later is partially confirmed. However, the hypothesis was based on two implicit premises: (I) that new forms of knowledge would be either highly similar or completely different from pre-existing ones; (ii) that knowledge variety could not be decomposed.

In the present paper we replace these premises by measures of three properties of the knowledge base of the KISs we studied and we introduce the distinction between intra- and inter-group knowledge variety. This more accurate and subtler approach leads to possible outcomes which could previously not have been envisaged. For example, we can find the combination of a growing trend in a given variable and of fluctuations in the same variable during a given period of time. These combinations can give rise to multiple outcomes such as a growing coherence when the new forms of knowledge have a very low cognitive distance with respect to the previous ones or to falling coherence for a much higher cognitive distance. Furthermore, the predominance of intra-group knowledge variety is likely to be associated with a lower cognitive distance and with a higher coherence than the predominance of inter-group knowledge variety. Although our hypotheses are broadly confirmed, the presence of these measures and of these distinctions will require the hypotheses to be refined.
Our results still confirm the importance of discontinuities in knowledge and allow us to maintain the fruitful distinction between random and organised search strategies. In

general we still expect that the higher the rate of increase over time in variety and in cognitive distance, the higher the decrease over time in coherence in the knowledge base, the more persistent the period of random screening, i.e. the less established the organized screening period. Both in biotechnology and telecommunications, the organized screening period seems to be more established, while the evolution of the KB in electronics is more characteristic of a persistent period of random screening. Yet the distinctions and measures we introduced in this paper give rise to some outcomes we could not have previously expected.

With our research we begin to make sense of the historical paths of development of these sectors. The biotechnology sector faced a major discontinuity in the knowledge base when the discoveries of recombinant DNA and monoclonal antibodies opened the door to a large range of industrial applications. The change was very important, since the knowledge base had previously been constituted mainly of organic chemistry. The replacement of a very large share of the existing research personnel by researchers with new and very different competencies occurred. Here, researchers working in public research institutes played a key role. A lot of them became scientific entrepreneurs and created their own businesses, often called the new dedicated biotechnology firms. A relationship of complementarity occurred between these new actors and the incumbent ones, the new actors providing new scientific and technological knowledge, while the incumbent ones developed complementary assets (finance, marketing, etc). A gradual improvement of the coherence of the knowledge base was achieved over time thanks to this relationship of complementarity often materialised by cooperation agreements between the two actors, although variety and cognitive distance in the new knowledge remain significant.

A similar process occurred in the telecommunications industry. The emergence of packet-switched technologies on which the Internet is based generated a new set of commercial applications. The former knowledge base in the telecommunications sector was essentially related to the circuit-switched technology, and required a drastic change in competencies to adapt to the new industrial challenges. Here also, new technology based firms (IP based) were created and contributed to the gradual change in the knowledge base, evolving towards a greater coherence. The process of merger and acquisitions between incumbent and new actors has certainly been more pronounced than in the biotechnology sector, however.

The electronics sector proceeded a different way. Part of the explanation can be that the sector is very vast, involving that it may be difficult to find the general explanation of the evolution of the knowledge base of this sector. However, it may be conjectured that since the most important IPC fields in the sector are common to the telecommunications sector (HO4M, telephonic communication), or highly related to it (HO5K, printed circuits), the evolution of the knowledge base can progressively become more comparable with the one observed in the telecommunications sector, although with an important delay.

7. Conclusion

This paper was intended to characterize the KB of three different KISs, namely biotechnology, telecommunications and electronics. Several major striking features emerge from these results. The first one is that biotechnology, telecommunications and electronics can effectively be qualified as KISs compared to other sectors. The second one is that all of these sectors, when confronted to the emergence of a discontinuity, proceed from a period of random screening to a following period of more organized search. The characteristics of the KB for each of these KISs differ from one period to the other, and attest the transformation of the KB before and after the emergence of the discontinuity.

References:

Archontopoulos, E., Guellec, D., Stevnsborg, N., Van Pottelsberghe de la Potterie, B., Van Zeebroeck, N., 2007, "When Small Is Beautiful: Measuring the Evolution and Consequences of the Voluminosity of Patent Applications at the EPO", *Information Economics and Policy*, 19(2), 103-32.

Breschi, S., Lissoni, F., and Malerba, F., 2003, "Knowledge relatedness in firm technological diversification", *Research Policy*, 32, 69-97.

Grebel, T., Krafft, J., Saviotti, P.P., 2006, "On the Life Cycle of Knowledge Intensive Sectors", *Revue de L'OFCE*, june, 63-85

Krafft, J., 2004, "Entry, Exit and Knowledge: Evidence from a Cluster in the Info-communications Industry, *Research Policy*, 33(10), 1687-1706

Nesta, L., 2007, "Knowledge and productivity in the world's largest manufacturing corporations", *Journal of Economic Behavior and Organization*, article in press.

Nesta, L., and Saviotti, P.P., 2005, "Coherence of the Knowledge Base and the Firm's Innovative Performance: Evidence from the U.S. Pharmaceutical Industry", *Journal of Industrial Economics*, 53(1), 123-42

Nesta, L., and Saviotti, P.P., 2006, "Firm Knowledge and Market Value in Biotechnology", *Industrial and Corporate Change*, 15(4), 625-52

Nooteboom B., 1999, *Inter-firm Alliances: Analysis and Design*, London, Routledge

Nooteboom B., 2000, Learning by Interaction: Absorptive Capacity, Cognitive Distance and Governance, Vol 4, N° 1-2, 69-92.

Penrose, E., 1959, *The Theory of the Growth of the Firm*, Basil Blackwell: Oxford.

Piscitello, L., 2004, "Corporate Diversification, Coherence and Economic Performance", *Industrial and Corporate Change*, 13(5), 757-87

Saviotti, P.P., 1996, Technological Evolution, Variety and the Economy, Cheltenham: Edward Elgar.

Saviotti P.P., Catherine D., 2008, "Innovation networks in biotechnology", forthcoming in *Handbook of Bioentrepreneurship*, Holger Patzelt, Thomas Brenner, David B. Audretsch (Eds) Springer.

Teece, D. et al., 1994, "Understanding Corporate Coherence: Theory and Evidence", *Journal of Economic Behavior and Organization*, 23(1), 1-30

Verspagen, B., 1997, "Measuring inter-sectoral technology spillovers: estimates for the European and US patent office databases, *Economic System Research*, 9(1), 49-67.